

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-155758

(P2000-155758A)

(43)公開日 平成12年6月6日(2000.6.6)

(51)Int.Cl.⁷

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/40

テーマコード(参考)

3 1 0 C 5 B 0 7 5

3 1 0 F

3 7 0 A

審査請求 未請求 請求項の数7 OL (全6頁)

(21)出願番号

特願平10-328940

(22)出願日

平成10年11月19日(1998.11.19)

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 岩山 真

埼玉県比企郡鳩山町赤沼2520番地 株式会
社日立製作所基礎研究所内

(72)発明者 西岡 真吾

埼玉県比企郡鳩山町赤沼2520番地 株式会
社日立製作所基礎研究所内

(74)代理人 100061893

弁理士 高橋 明夫 (外1名)

最終頁に続く

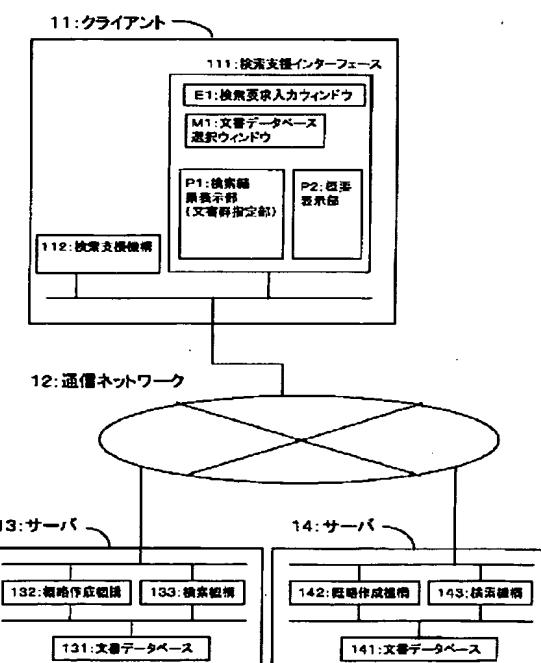
(54)【発明の名称】複数文書データベースを対象とした文書検索方法および文書検索サービス

(57)【要約】

【課題】複数の文書データベース間の関連性を調べる
ような文書検索を効率良く実現すること。

【解決手段】各文書データベースに概略作成機構と検
索機構を持たせ、それぞれサーバとして通信ネットワー
クに接続する。クライアントは指定した文書データベー
ス内の文書群から、その概略を介して、指定した文書デ
ータベース内の関連する文書群を得る。得られた概略を
他のサーバに送り、転送先のサーバの文書データベース
で転送された概略に応じた検索を行う。

図1



【特許請求の範囲】

【請求項 1】複数の文書データベースを検索対象として切りかえる機能およびインターフェースを有し、キーワード集合、あるいは文書の断片、あるいは任意の文書群を検索入力として受けつけ、検索入力と関連度（類似度）の高い文書群を、選択した文書データベースから関連の強さの順に検索する機能を有する文書検索方法であって、前記一つの文書データベースの検索結果を検索対象を切り替えて他のデータベースの検索の検索入力とすることができる特徴とする文書検索方法。

【請求項 2】一つの文書データベースの検索結果の文書群について選択／非選択が指定できるインターフェースを有し、選択された検索結果の文書群を、他の検索対象の検索の検索入力とする請求項 1 記載の文書検索方法。

【請求項 3】検索入力中の特徴的な単語のみを選択した概要を検索に利用する請求項 1 または請求項 2 記載の文書検索方法。

【請求項 4】各文書データベースとそれを操作するプログラムを有するサーバが通信ネットワーク上に分散して配置され、クライアントは選択した文書データベースが格納されているサーバに対して検索入力の文書群を送信し、前記サーバから送信した文書群に関する特徴的な単語のみを選択した概要を受信し、受信した概略に対する利用者の評価結果を反映した前記概要に対応する検索入力を他の文書データベースが格納されているサーバに送り、検索結果を受信する機能を有する請求項 1 または請求項 2 記載の文書検索方法。

【請求項 5】前記サーバはクライアントから送られてきた文書群に対して特徴的な単語からなる概略を作成してクライアントに送信し、クライアントから送られてきた任意の概略に対しては関連度の高い文書群を検索しクライアントに送信する請求項 4 記載の文書検索方法。

【請求項 6】前記クライアントは検索入力の文書群と検索先の文書データベースを指定するインターフェースを有し、検索入力の文書群を利用者が指定するサーバに送り、当該サーバからそれら文書群の概略を受信し、受信した概略を他の文書データベースを含むサーバに送り、当該サーバから検索結果を受信し表示する機能を有する請求項 4 記載の文書検索方法。

【請求項 7】通信ネットワーク上に分散して配置された各文書データベースとそれを操作するプログラムを有するサーバを保有し、該サーバに接続されたクライアントに文書検索をサービスする文書検索サービスであって、前記クライアントは選択した文書データベースが格納されているサーバに対して検索入力の文書群を送信し、前記サーバから送信した文書群に関する特徴的な単語のみを選択した概要を受信し、受信した概略に対する利用者の評価結果を反映した前記概要に対応する検索入力を他の文書データベースが格納されているサーバに送り、検

索結果を受信する文書検索方法を可能とするとともに、前記サーバはクライアントから送られてきた文書群に対しては特徴的な単語からなる概略を作成してクライアントに送信し、クライアントから送られてきた任意の概略に対しては関連度の高い文書群を検索しクライアントに送信する文書検索サービス。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、複数の文書データベースを切りかえて検索したり、複数の文書データベースを関連付けたりするための文書検索方法に関する。

【0002】

【従来の技術】様々な文書情報が電子化されるにつれ、複数の異質な文書データベースを同時に検索する必要性が増してきている。例えば、興味を引いた新聞記事からそれに関連する百科辞典の項目を閲覧するといった要求は多い。

【0003】従来の検索技術においても複数の文書データベースを切りかえて検索することは可能であるが、ある文書データベース内の文書群に対して、別の文書データベース内の関連する文書群を検索するといった、文書データベース間に渡る文書群間の関連性を調べることはできなかった。

【0004】これに対し、同一文書データベースに限れば、文書群を検索入力として指定して、同じ文書データベース内の関連する文書群を検索することは可能である。この場合、検索に先立って文書間の関連度を計算しておくことにより十分な検索速度を得る場合が多い。異なる文書データベース間においても、このような前計算を行えば、複数文書データベースを同時に検索することも可能であるが、文書データベースの数が増すにつれ、前計算の必要数も組み合わせ的に増大するため、この方法も現実的には不可能である。

【0005】また、利用者側で一旦検索元の文書群を解析し、検索入力を構成して、他の文書データベース内を検索することも可能であるが、この場合、利用者側が検索元の文書群に関する全情報を受けとらなければならず、文書データベースが通信ネットワーク上に存在する場合、通信量が膨大になってしまう。

【0006】

【発明が解決しようとする課題】前記従来技術の問題を解消し、利用者が、任意の文書データベース中の任意の文書群を指定し、その文書群に関連する文書群を、更に任意の文書データベース内から効率よく検索できるようすることである。

【0007】

【課題を解決するための手段】文書群のように検索入力が大きい場合、検索入力の全情報を検索に使うのではなく、検索入力内の特徴的な単語のみを概要として検索に使うことで、検索速度が速く、かつ、通信ネットワーク

への負荷も小さい検索方法を実現する。

【0008】各文書データベースに関しては、指定された文書群に対して、その中で特徴的な単語を選択することで概要を作成する機構と、送られてくる任意の概要に対して検索を行う機構を有するサーバーとして通信ネットワークに配置する。

【0009】検索を行う利用者は、クライアントを介して、まず、検索元の文書データベースが格納されているサーバーに対して文書群を指定して、その概要を受けとる。次に受けとった概要を検索先の文書データベースが格納されているサーバーに送り、検索結果を受けとる。

【0010】クライアントの検索インターフェースとしては、まず、文書群の表示エリアを有し、このエリアにおいて必要な文書群を指定することができるようになる。また検索先のデータベースも選択できるようになる。これによりクライアントでは、文書群表示エリアに表示されている文書群の中から、利用者が興味ある文書群を選択して、必要なら検索先の文書データベースを切りかえて検索を行うことができる。

【0011】

【発明の実施の形態】図1は、クライアント11がサーバ13の文書データベース131内の任意の文書群を指定して、指定した文書群と関連度（類似度）の高い文書群を別のサーバ14の文書データベース141から得るための方法を実現する全体構成の一例を示したものである。ここで、検索元、検索先の文書データベース131、141は通信ネットワーク12を介してアクセスできる異なった場所にあるサーバ上にそれぞれ配置されている。

【0012】まず、クライアント11は、利用者の入力に応じて検索元となる文書データベース131内の文書群を指定し、サーバ13が理解できる文書識別子の集合として通信ネットワーク12を介してサーバ13に送出する。文書群の指定は、後述する検索結果表示部（文書群指定部）P1にて行う。

【0013】サーバ13は、検索機構133により、クライアントから送られてきた文書識別子の集合と関連度の高い文書群を文書データベース131から検索する。この際、概略作成機構132により、検索された文書群に対して文書群の概略を作成し、通信ネットワーク12を介してクライアント11に返答する。ここでの概略とは、文書群をよくあらわす単語の集合のことある。概略作成機構の実施形態は特開平9-62693「確率モデルによる文書分類方法」などの既存の方法が利用できる。

【0014】一例を示すと、まず、概略を作成しようとする文書群中の全文書を単語に分割して頻度集計する。一般に、ある文書群で良く現れる単語ほどその文書群を代表する度合も高いため、文書群中で出現頻度が高い単語ほど概略に含まれやすいことになる。ただし「する」

などのように、どの文書にもよく現れるような一般語は概略として適当ではない。よって通常は、文書群が属する文書データベース中での出現頻度も考慮して概略としての単語選択を行う。つまり、指定された文書群中での出現頻度が高く、かつ、文書データベース全体での総出現頻度が低い単語ほど、その文書群でしか現れないという意味で特徴的な単語であり、その文書群を特徴付ける概略として適切である。具体的には、文書群中のそれぞれの単語について、文書群中での出現頻度、文書データベース中での出現頻度を入力とする適当な関数により単語の重みを計算し、ある閾値以上の重みを持つ単語を概略として採用する。サーバ13は以上的方法で作成した重み付き単語の集合を通信ネットワーク12を介してクライアントに返す。この単語を図2では、「特徴語」として表示するものとしている。

【0015】次に、クライアント11はサーバ13から返答された概略（検索元の文書群の概略）を評価しあるいは加工して、通信ネットワーク12を介して検索先のサーバ14に送出する。

【0016】サーバ14は、検索機構143により、クライアントから送られてきた文書群の概略と関連度の高い文書群を検索先の文書データベース141から検索し、検索結果の文書識別子を関連度の重み付きでクライアント11に返す。ここでの検索機構は、公知のキーワード検索法により実現できる。つまり、入力である文書群概略は重み付き単語の集合であるため、各単語を重み付きの入力キーワードとみなしちOR検索すればよい。その際、検索結果の文書の重み（関連度）は以下のように計算できる。概略および検索先の文書両方に含まれる各単語について、概略中の重みと、検索先の文書における重み（例えば頻度）から総合的な重みを計算し（例えば両重みの積）、さらにそのような単語全てに関する重みを集計（例えば総和）することで関連度を得る。

【0017】以上的方法で、クライアント11は文書データベース131内の任意の文書群に関する文書データベース141内の文書群を得ることができる。ここでの特徴は、検索元の文書群に関する処理（概略作成）をサーバ側に任せることにより、通信ネットワーク中の通信量が少なくて済む点である。クライアントが検索元の文書全文情報を一旦受けとめて処理する場合に比べると差は歴然である。クライアントの検索支援機構112では、基本的には検索元のサーバから送られてきた文書群概略を検索先サーバに送るだけによく、検索に関わる処理のほとんどを両サーバにまかせることができる。一方サーバ側は、担当する文書データベースに関して、概略作成機構、検索機構を持つのみでよく、モジュール化されており、他の文書データベースの情報に関しては全く考慮しなくてもよい。

【0018】以上、文書データベース131を検索元として文書データベース141を検索する手段を説明した

が、全く同様の方法で、逆に文書データベース141を検索元として文書データベース131を検索することも可能である。この場合、クライアントは、文書データベース141中の文書群の概略をサーバ14の概略作成機構142から得て、検索先のサーバ13に送信し、サーバ13の検索機構133により文書データベース131中の関連する文書群を得る。以上を一般化すると、新たな文書データベースに関しては、その概略作成機構、検索機構を持つサーバを用意し、通信ネットワークに接続するだけで、その文書データベースは、通信ネットワークに接続されている全ての文書データベースに対して検索元にも検索先にもなり得る。

【0019】最後にクライアントに関する実施形態を図2で説明する。111はクライアントに搭載されている検索支援インターフェースの例で、これは基本的には、特願平9-240963「文書検索支援方法および文書検索支援サービス」で本願の発明者らによって提案されたものと同じである。E1は検索要求の入力ウィンドウであり、利用者は、ここに検索要求をキーワードの羅列、または文章形式で入力できる。M1は文書データベース選択ウィンドウであり、利用者が、右端の指示部をマウスでプルダウンすることにより文書データベースの一覧があらわれ、所望の文書データベースを選択できる。B1は検索の開始を指示するボタンである。よって、利用者はウィンドウE1に任意の検索要求を入力し、ウィンドウM1で検索対象の文書データベースを選択し、ボタンB1を押すことで、ウィンドウM1で選択した文書データベースに対してウィンドウE1に入力したキーワードによる通常のキーワード検索を実行させることができる。この検索の実行は図1に示す検索支援機構112の支援のもとに実行されるが、これの詳細は、先の出願に詳しいので、ここでは説明を省略する。もちろん、一般に行われるキーワード検索によっても良いことは言うまでもない。

【0020】P1は検索結果表示部であり、上段に選択の結果選択された文書の総数および後述するようにして利用者に選択された文書の数を表示する窓が、その下に、利用者の選択／非選択を入力する窓、検索要求との関連度および選択された文書のタイトルの一覧（リスト形式で表示）を示す表示部が配置されたものとなっている。この表示部はスクロール機能を持っており表示に一度に表示できない部分もスクロールによって見ることができる。選択／非選択を入力する窓はマウスによりクリック可能であり、マウスでクリックする毎に選択／非選択の状態が反転する。これをクリックして選択すると、この文書に対応した文書の概要が、概要表示部P2に重み付きの単語集合のグラフ形式で表示される。概要表示部P2にも、上段部に、特徴語の総数および利用者に選択された特徴語の数を表示する窓が設けられる。なお、文書タイトルは、通常、関連度の順にソートされて

いる。

【0021】図の検索結果表示部P1は、検索の結果選択された文書が総数で22あり、利用者が選択された文書のタイトルから興味ある文書として三つの文書を選択を選択している状態を示す。選択された文書はクリックによってチェックマークが表示されている。概要表示部P2は、これに応じて、選択された文書の検索要求入力に対応する特徴語が5つ表示されている。

【0022】この実施例では説明を省略するが、概要表示部P2に表示された特徴語を選択することでこれをキーとする文書を検索結果表示部P1に逆に表示させることもできる。したがって、利用者が自分の好みにカスタマイズした概要によって、よりきめ細かい検索が可能になる。これについては、先に引用した特願平9-240963に詳しく説明されている。

【0023】利用者が、このように、選択された文書のタイトルを参照しながら、文書の選択／非選択を行い、興味のある文書を複数個選択することができる。

【0024】その後、利用者が、この検索結果に対応した文書群について他の文書データでの扱い等に興味を持ったときは、ウィンドウM1により文書データベースを切りかえて、検索の開始を指示するボタンB1を押す。

【0025】クライアントは、これに応じて、検索元の文書データベースが格納されているサーバ（例えばサーバ13）に選択した複数文書の識別子を送り、それら複数文書の概要を得て、検索先の文書データベースが格納されているサーバ（例えばサーバ14）にこの概要を送り、検索先のサーバ（例えばサーバ14）から検索結果を得る。新しい検索結果は検索結果表示部P1に表示される。つまりこの例の場合、P1は新しく検索された文書群に書きかえられる。

【0026】新しい検索結果と、先の検索結果とを比較するため、先の検索結果をふたたび検索結果表示部P1に表示したければ、利用者はボタンB2を押して、検索結果表示部P1の表示を検索前の状態に戻すことができる。同様にボタンB3を押して、検索結果表示部P1の表示を新しい検索結果に進めることができる。

【0027】このような検索結果に応じた他の文書データベースでの検索は、検索の任意の段階で実行することができるから、検索サイクルを繰り返すことにより利用者は文書データベースから文書データベースへと探索を自由に進めることができる。当然、文書データベースを切りかえずに同一文書データベース内でこのサイクルを繰り返すことも可能である。

【0028】

【発明の効果】利用者は、各文書データベースの配置、構成について意識すること無く、検索対象としての文書データベースを自由に指定して、探索を自由に進めることができる。また、文書データベースを保有するサーバ

はモジュール化できるため、新たな文書データベースに関するでは、その概要作成機構、検索機構を有するサーバを通じてネットワークに接続するだけで、通信ネットワーク上の他の文書データベース全てに対して、検索先にも検索元にもなることが可能である。

【図面の簡単な説明】

【図1】複数文書データベース検索方法のシステムの全体構成の一例を示す図。

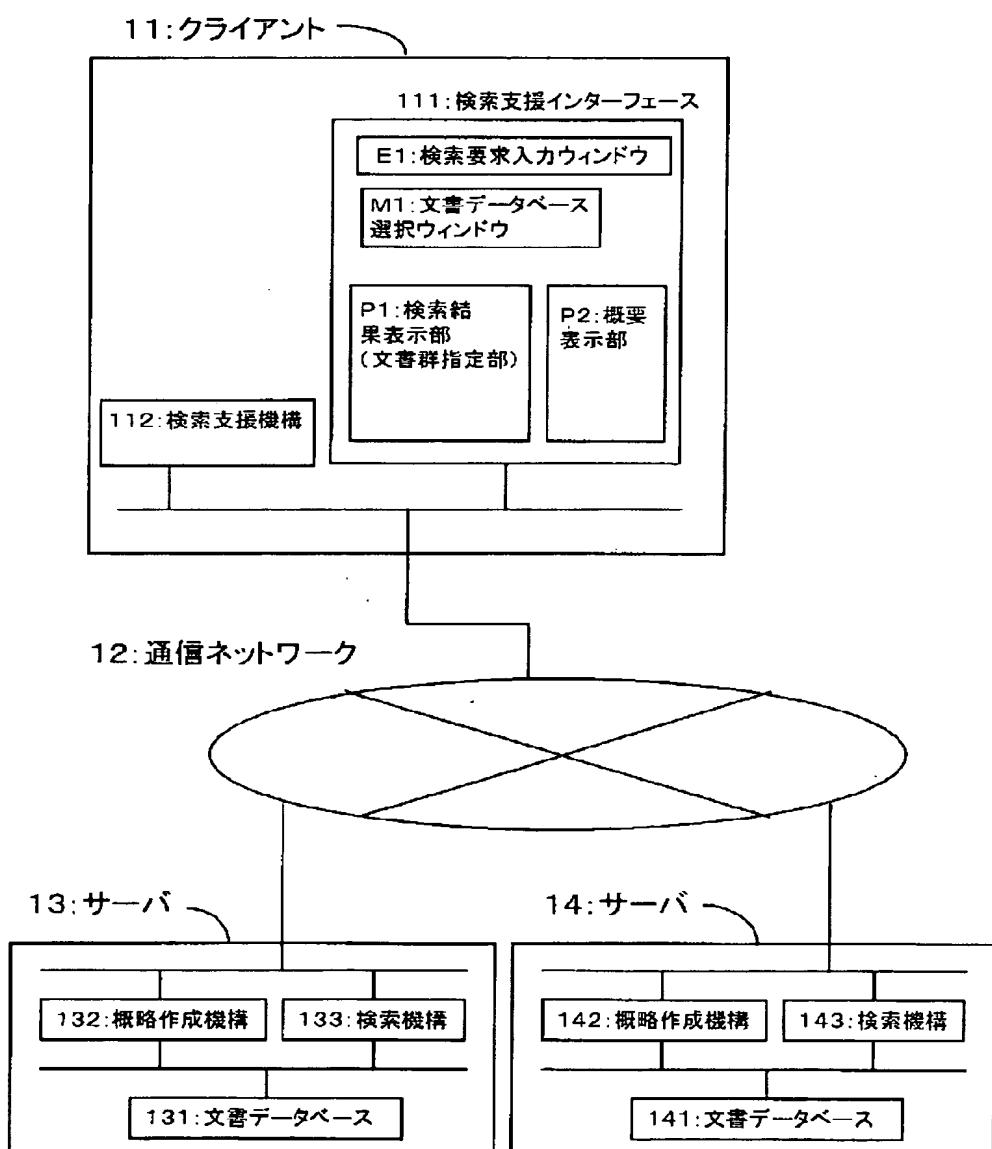
【図2】クライアントにおける検索支援インターフェースの構成の一例を示す図。

【符号の説明】

11: クライアント、111: 検索支援インターフェース、112: 検索支援機構、12: 通信ネットワーク、13: サーバ、131: 文書データベース、132: 概略作成機構、133: 検索機構、14: サーバ、141: 文書データベース、142: 概略作成機構、143: 検索機構、144: 検索結果表示部、B1: 検索ボタン、B2: 戻りボタン、B3: 進むボタン、E1: 検索要求入力ウインドウ、M1: 文書データベース選択ウインドウ、P1: 検索結果表示部、P2: 概要表示部。

【図1】

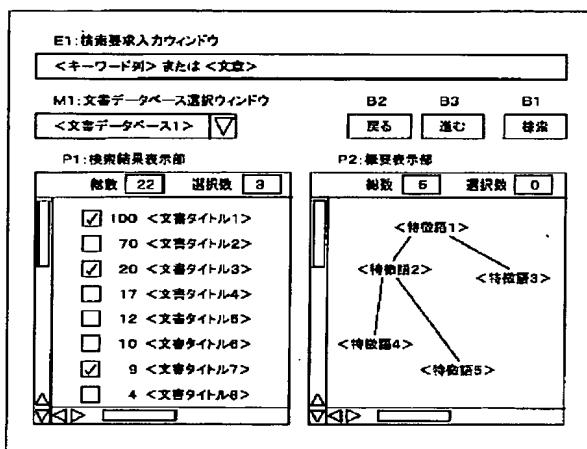
図1



【図2】

図2

111:検索支援インターフェース



フロントページの続き

(72)発明者 丹羽 芳樹

埼玉県比企郡鳩山町赤沼2520番地 株式会
社日立製作所基礎研究所内

(72)発明者 高野 明彦

埼玉県比企郡鳩山町赤沼2520番地 株式会
社日立製作所基礎研究所内

F ターム(参考) 5B075 KK03 KK07 ND03 NS10 PP02
PP03 PP12 PP13 PP22 PP24
PP30 PQ02 PQ05 PQ40 PQ46
PQ66 PQ75 UU06